Automatica 93 (2018) 476–488

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

# Guaranteed model-based fault detection in cyber–physical systems: A model invalidation approach[☆]

Farshad Harirchi [*], Necmiye Ozay

*Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109, United States*

ARTICLE INFO

*Article history:*
Received 4 October 2016
Received in revised form 4 February 2018
Accepted 6 March 2018

ABSTRACT

This paper presents a sound and complete fault detection approach for cyber–physical systems represented by hidden-mode switched affine models with time varying parametric uncertainty. The fault detection approach builds upon techniques from model invalidation. In particular, a set-membership approach is taken where the noisy input–output data is compared to the set of behaviors of a nominal model. As we show, this set-membership check can be reduced to the feasibility of a mixed-integer linear programming (MILP) problem, which can be solved efficiently by leveraging the state-of-the-art MILP solvers. In the second part of the paper, given a system model and a fault model, the concept of $T$-detectability is introduced. If a pair of system and fault models satisfies $T$-detectability property for a finite $T$, this allows the model invalidation algorithm to be implemented in a receding horizon manner, without compromising detection guarantees. In addition, the concept of weak-detectability is introduced which extends the proposed approach to a more expressive class of fault models that capture language constraints on the mode sequences. Finally, the efficiency of the approach is illustrated with numerical examples motivated by smart building radiant systems.

## 1. Introduction

cyber–physical systems are combinations of physical processes and embedded computers. The embedded computers collect data from the process through sensors and control it in a closed-loop manner. With the increase in data acquisition and storage capacity and the decrease in sensor costs, it is possible to collect large amounts of data during the operation of complex cyber–physical systems. For instance, "a four-engine jumbo jet can create 640 TB of data in just one crossing of the Atlantic Ocean" (Rajah, 2014). As discussed in Sznaier, Camps, Ozay, and Lagoa (2014), this exponential growth in the data collection capabilities is a major challenge for systems and control community. Sensor/information-rich networked cyber–physical systems, from air traffic or energy networks to smart buildings, are getting tightly integrated into our daily lives. As such, their safety-criticality increases. For such systems, it is crucial to detect faults or anomalies in real-time to support the decision-making process and to prevent potential large-scale failures.

### 1.1. Contributions

This paper presents a fault detection scheme to enhance the reliability of a class of cyber–physical systems that are represented by hidden-mode switched affine models with time-varying parametric uncertainty. This modeling framework is quite expressive and can be used to describe a wide range of cyber–physical systems such as heating, ventilation and air conditioning (HVAC) systems in smart buildings (Weimer et al., 2013), wind turbines (Burkart, Margellos, & Lygeros, 2011), power systems and power electronics, automotive systems, aircraft, air traffic, and network and congestion (Sun, 2006). We model faults also in this framework allowing us to capture many scenarios including cascaded faults or various types of cyber or physical attacks. Note that linear time invariant systems with or without noise or affine parametric uncertainty are special cases of the modeling framework.

The proposed fault detection scheme builds on set-membership model invalidation approaches (Ozay, Sznaier, & Lagoa, 2010, 2014; Rosa, Silvestre, Shamma, & Athans, 2010). Unlike many set-membership methods that compute explicit set representations and propagate them via set-valued observers, the proposed

[☆] This work is supported in part by DARPA grant N66001-14-1-4045 and an Early Career Faculty grant from NASAs Space Technology Research Grants Program. The material in this paper was partially presented at the 5th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS), October 14–16, 2015, Atlanta, GA. This paper was recommended for publication in revised form by Associate Editor Michele Basseville under the direction of Editor Torsten Söderström.

[*] Corresponding author.
*E-mail addresses:* harirchi@umich.edu (F. Harirchi), necmiye@umich.edu (N. Ozay).

https://doi.org/10.1016/j.automatica.2018.03.040
0005-1098/© 2018 Elsevier Ltd. All rights reserved.

method uses an online optimization formulation for model invalidation that keeps implicit constraints to represent sets. In particular, we show that model invalidation problem for this class of systems can be reduced to the feasibility of a MILP problem, which can be checked efficiently using state-of-the-art solvers (CPLEX, 2009). Additionally, the concept of $T$-detectability is introduced, which enables us to apply the model invalidation approach for fault detection in a receding horizon manner (with a horizon size $T$) without losing detection guarantees. Even though there are some practical systems with fault models that are $T$-detectable, a limited class of system and fault models satisfy this property. We further discuss weak detectability that incorporates language constraints on the switching sequences of the faults to enable detection of a broader class of faults. Algorithms that can be used to find the minimum $T$ (if it exists) are presented.

A preliminary version of this paper is published in Harirchi and Ozay (2015). The current paper significantly extends the model class by allowing uncertainty in variables and considering language constraints in the mode sequence. Moreover, more efficient MILP-based necessary and sufficient conditions for verifying $T$-detectability are given and some connections to mode observability of switched systems are pointed out.

### *1.2. Literature review on fault detection*

Model-based fault detection has a long history starting with early work on failure detection filters (Beard, 1971; Jones, 1973). A vast majority of fault detection methods are based on residual generation, where the residual is evaluated by simple thresholding methods or more complicated classifiers to decide between faulty and normal behaviors (Ding, 2008; Isermann, 2006; Patton, Frank, & Clark, 2013; Simani, Fantuzzi, & Patton, 2003). The residual generation methods are classified into three main categories (Simani et al., 2003): parameter estimation-based (Isermann, 1993), observer-based (Frank, 1993; Patton & Chen, 1997; Shames, Teixeira, Sandberg, & Johansson, 2011) and parity equation-based (Gertler, 1997) techniques. All these residual generation techniques can be implemented in real-time, but even when a specific fault model exists, their behavior is usually analyzed only asymptotically and they fail to provide any finite-time detection guarantees.

As an alternative to residual generation, set-membership fault detection methods have been proposed both for passive (Rosa & Silvestre, 2013; Rosa et al., 2010) and active (Nikoukhah, 1998; Nikoukhah & Campbell, 2006; Scott, Findeisen, Braatz, & Raimondo, 2014) fault detection. Some of these methods proceed by computing convex-hulls of potentially non-convex reachable sets of the system and comparing the actual output to this set (Rosa et al., 2010). Since, reachable sets are over-approximated, this leads to only sufficient conditions, that is, they guarantee there are no false alarms. Scott et al. (2014) pose a mixed-integer quadratic programming problem to find an optimal separating input to detect faults. They use zonotopes to represent and propagate state constraints. The order of the zonotopes is used to trade-off between the complexity and the conservativeness of the approach. This approach is extended in Raimondo, Marseglia, Braatz, and Scott (2016) to handle admissible sets in the form of constrained zonotopes and to take advantage of real-time measurements to calculate an improved separating input. Despite the fact that all the above mentioned set-membership methods are proposed for linear systems, their scalability is somewhat limited to be applied in real-time (Scott et al., 2014).

Non-linear and hybrid systems have also attracted notable attention from fault detection community. Most of the research is concentrated around residual generation type methods (Garcia & Frank, 1997; Hammouri, Kinnaert, & El Yaagoubi, 1999; Pan, Yuan, Sandberg, Gonçalves, & Stan, 2015). De Persis and Isidori (2001) develop analytical necessary and sufficient conditions under which the problem of fault detection and isolation becomes solvable for non-linear systems. However, they do not particularly address the computational aspects of the problem. Observer-based methods are employed for fault diagnosis in hybrid systems both when the discrete mode is observed (McIlraith, Biswas, Clancy, & Gupta, 2000) or hidden (Narasimhan & Biswas, 2007; Wang et al., 2018), using variants of Kalman filters. In recent work (Deng, D'Innocenzo, Di Benedetto, Di Gennaro, & Julius, 2016), Deng et al. investigate fault diagnosis problem in hybrid systems, by constructing a finite abstraction for the hybrid automaton and analyzing the diagnosability of the abstract system using tools from discrete-event systems, providing some detection guarantees. However, the mode signal is assumed to be observed in Deng et al. (2016).

The notion of $T$-detectability is closely related to distinguishability of dynamical systems (De Santis & Di Benedetto, 2016; Grewal & Glover, 1976) and observability in switched systems (Babaali & Egerstedt, 2004; Vidal, Chiuso, & Soatto, 2002). If two systems are distinguishable, then there exists a non-zero input which makes their trajectories different at least for one time instance. In De Santis (2011), it is illustrated that if two models are distinguishable for a particular control input, then they are distinguishable for a generic control law. Lou and Si (2009) introduced the concept of input-distinguishability, which restricts distinguishability to all non-zero inputs contained in a convex set. Rosa and Silvestre (2011) consider the input-distinguishability for discrete time linear time-invariant systems with bounded disturbance and noise. They provide a necessary and sufficient rank condition to check for a given size of time horizon ($T$) if two systems are input distinguishable or not. This condition is seldom satisfied in practice, hence they added extra constraint which enforces the persistence of excitation for disturbance (Rosa & Silvestre, 2011). The concept of $T$-detectability introduced here is closely related to absolute input-distinguishability (Rosa & Silvestre, 2011). In other words, $T$-detectability of a pair of hidden-mode switched affine models subject to process and measurement noise provides necessary and sufficient conditions for their input-distinguishability. The time horizon $T$ in detectability is an upper bound on the detection delays (time from the occurrence of fault to detection alarm) (Adnan, Izadi, & Chen, 2011; Mariton, 1989; Stoorvogel, Niemann, & Saberi, 2001).

Model invalidation was originally proposed as a way to build trust in models obtained through a system identification step or discard/improve them before using these models in robust control design (Smith & Doyle, 1992). In model invalidation problem, one starts with a family of models (i.e., a priori or admissible model set) and experimental input–output data collected from a system (i.e., a finite execution trace) and tries to determine whether the experimental data can be generated by one of the models in the initial model family. Its relation to fault and anomaly detection has been pointed out in Cheng, Wang, Sznaier, Ozay, and Lagoa (2012), Ozay et al. (2010), Ozay et al. (2014) and Rosa et al. (2010) for linear time-varying systems and hybrid systems in autoregressive form. A fault detection scheme based on model-invalidation for polynomial state space models subject to noise and uncertainty in parameters is recently proposed in Harirchi, Luo, and Ozay (2016). The convex relaxations for model invalidation problem mostly provide sufficient conditions that can be efficiently checked to detect faults but only necessary for large relaxation orders.

The rest of the paper is structured as follows. In Section 2, we describe the class of systems that are considered in this work. The model invalidation and $T$-detectability problems are defined in Section 3. The main contributions of the paper including the proposed fault detection scheme are presented in Section 4. Section 5 describes the generalizations of the proposed framework for